

В. И. БУЛКИН, канд. техн. наук, доц. Макеевского экономико-гуманитарного института (г. Макеевка)

О ПРЕДСТАВЛЕНИИ ЗНАНИЙ В ИНТЕЛЛЕКТУАЛЬНЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

У статті розглянуто питання щодо множин однорідних і комплексних форм подання семантичної інформації. Показано, що множини однорідних й комплексних форм представлення інформації можуть бути задані у вигляді рівнянь алгебри скінчених предикатів. Одержані рівняння припускають схему реалізації у вигляді перемикаючих ланцюгів.

Актуальность. Изобретение современных вычислительных машин привело к настоящей революции в сфере информационных технологий. На базе использования высокопроизводительной компьютерной техники появились новые информационные технологии и высокие информационные технологии. Однако современные компьютеры все еще продолжают использоваться в качестве средства усиления интеллектуальных способностей человека. Современный компьютер может решать только формализуемые задачи, т. е. задачи, для решения которых можно составить алгоритм.

Решение сложных неформализуемых задач, задач творческого, эвристического характера является прерогативой человека. Для решения этих задач человеку необходим партнер. В качестве такого партнера могут выступать компьютеры будущих поколений, которые должны отличаться от современных компьютеров, прежде всего более высоким уровнем интеллекта, а также как можно более широкими знаниями об окружающей действительности. В идеале компьютер должен стать настоящим помощником человека в решении задач творческого характера, задач с неполными и нечеткими исходными данными и других задач, не поддающихся формализации. Для решения таких задач необходимо повышать уровень «интеллекта» компьютера. На пути решения проблемы интеллектуализации компьютера возникает немало трудностей, связанных с созданием дружественного интерфейса, использующего естественный язык для общения с пользователем.

Большие трудности возникают при попытке создания систем параллельной обработки информации, так называемых мозгоподобных компьютеров. Актуальной остается проблема создания систем понимания, обрабатывающих семантическую информацию. Предварительным этапом, который может помочь решению выше перечисленных задач, должен стать этап разработки и исследования функционирования информационных семантических систем, построенных на качественно ином принципе.

Информационные семантические системы – это системы, которые служат для достижения цели «понимания» естественно языковых запросов и адекватного реагирования на них. Главной особенностью этих систем является семантическая (смысловая, осмысленная) обработка семантической информации. Семантической информацией считается множество сведений о различных сторонах и отношениях изучаемых объектов заданной предметной области. Определение понятия семантической информации во многом совпадает с определением понятия знаний. Поэтому в дальнейшем, наряду с термином «семантическая информация» будем использовать термин «знания». Под семантической обработкой информации подразумевается:

- создание первичной семантической информации (первичных знаний) о заданной предметной области;
- преобразование одной формы представления первичной семантической информации (первичных знаний) в другую форму представления;
- преобразование первичной семантической информации (первичных знаний) во вторичную (сжатие первичной семантической информации);
- поиск семантической информации;
- передача семантической информации (знаний) от одного объекта к другому.

Постановка задачи исследования. В данной статье ставится задача описания различных форм представления знаний. Множества форм представления знаний предлагается описывать с помощью уравнений алгебры конечных предикатов, которые допускают их аппаратную реализацию в виде переключаемых схем. Определенный интерес представляет проблема преобразования первичной информации во вторичную и использование принципа семантической топологии для создания систем преобразования семантической информации из одной формы в другую.

Решение поставленных задач. Формы представления знаний в интеллектуальных информационных системах подразделяются на однородные и комплексные. Множество однородных форм представления знаний можно определить следующим выражением:

$$M_1 = \{t, s, v, g\}, \quad (1)$$

где t – текстовая форма представления знаний (книги, газеты, журналы, и т.д.), s – аудиальная форма представления знаний (речь, звуки, музыка), v – визуальная форма представления знаний (анимация, балет, пантомима, жесты, мимика), g – графическая форма представления знаний (картины, рисунки, схемы, графики и т. д.).

На языке алгебры конечных предикатов (АКП) множество M_1 можно записать в виде следующего уравнения:

$$x^t \vee x^s \vee x^v \vee x^g = 1, \quad (2)$$

где x – предметная переменная однородных форм представления знаний [1].

Очевидно, что совокупность корней данного уравнения совпадает с множеством M_1 . Переключательная цепь, реализующая уравнение (1), представлена на рис. 1 [2].

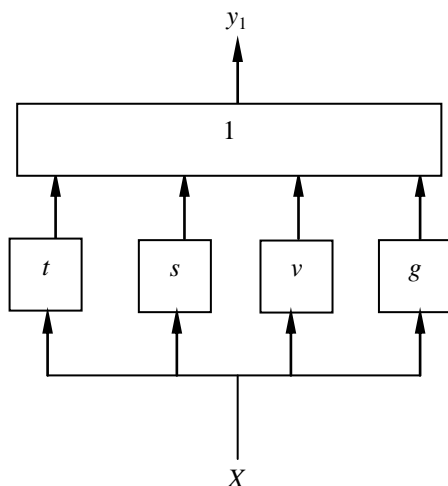


Рис. 1 Переключательная цепь

Комплексные формы представления знаний являются комбинацией однородных форм. Декартовым квадратом M_1^2 множества M_1 можно задать множество M_2 бинарных комплексных форм представления знаний:

$$M_2 = M_1^2 = \left\{ \begin{pmatrix} t \\ v \end{pmatrix}, \begin{pmatrix} t \\ v \end{pmatrix}, \begin{pmatrix} t \\ v \end{pmatrix}, \begin{pmatrix} t \\ v \end{pmatrix}, \begin{pmatrix} s \\ v \end{pmatrix}, \begin{pmatrix} s \\ v \end{pmatrix}, \begin{pmatrix} s \\ v \end{pmatrix}, \begin{pmatrix} s \\ v \end{pmatrix}, \begin{pmatrix} v \\ g \end{pmatrix}, \begin{pmatrix} v \\ g \end{pmatrix}, \begin{pmatrix} v \\ g \end{pmatrix}, \begin{pmatrix} v \\ g \end{pmatrix}, \begin{pmatrix} g \\ t \end{pmatrix}, \begin{pmatrix} g \\ s \end{pmatrix}, \begin{pmatrix} g \\ v \end{pmatrix}, \begin{pmatrix} g \\ g \end{pmatrix} \right\} \quad (3)$$

На языке алгебры конечных предикатов множество M_2 может быть задано следующим уравнением:

$$\begin{aligned}
& x_1^t x_2^t \vee x_1^t x_2^s \vee x_1^t x_2^v \vee x_1^t x_2^g \vee x_1^s x_2^t \vee x_1^s x_2^s \vee x_1^s x_2^v \vee x_1^s x_2^g \vee \\
& x_1^v x_2^t \vee x_1^v x_2^s \vee x_1^v x_2^v \vee x_1^v x_2^g \vee x_1^g x_2^t \vee x_1^g x_2^s \vee x_1^g x_2^v \vee x_1^g x_2^g = 1
\end{aligned} \tag{4}$$

где x_1, x_2 – переменные однородные формы представления знаний. Совокупность корней уравнения (4) совпадает с множеством M_2 бинарных комплексных форм представления знаний. Переключательная цепь, реализующая уравнение (4), представлена на рис. 2.

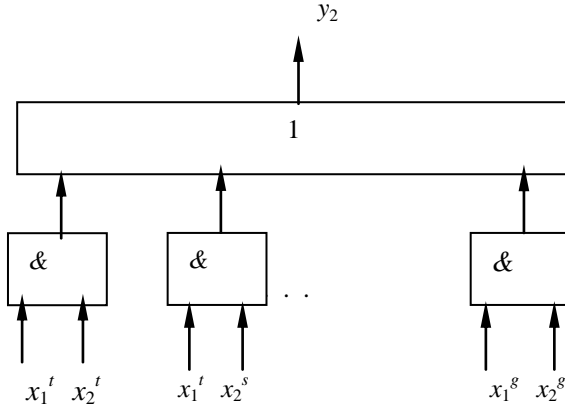


Рис. 2 Переключательная цепь, заданная для (4)

Декартов куб, множества M_1 , задает множество тернарных комплексных форм представления знаний и может быть записан следующим образом:

$$M_3 = M_1^3 = \{(t, t, t), (t, t, s), (t, t, v), (t, t, g), \dots, (g, g, t), (g, g, s), (g, g, v), (g, g, g)\} \tag{5}$$

Множество M_3 можно формально задать с помощью соответствующего уравнения алгебры конечных предикатов.

$$\begin{aligned}
& x_1^t x_2^t x_3^t \vee x_1^t x_2^t x_3^s \vee x_1^t x_2^t x_3^v \vee x_1^t x_2^t x_3^g \vee \dots \\
& x_1^g x_2^g x_3^t \vee x_1^g x_2^g x_3^s \vee x_1^g x_2^g x_3^v \vee x_1^g x_2^g x_3^g = 1
\end{aligned} \tag{6}$$

Корнями этого уравнения являются элементы множества M_3 . Уравнение (6) можно представить с помощью переключателя цепи (рис.3).

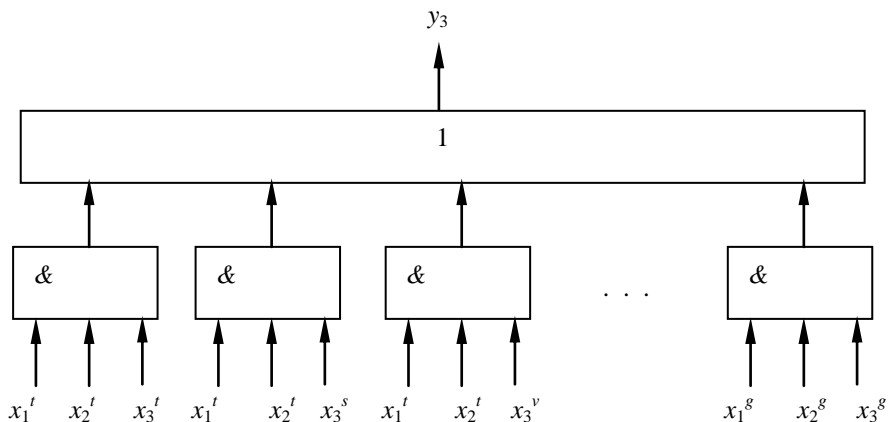


Рис. 3

Поступая аналогичным образом, можно получить множество комплексных форм представления знаний, состоящее из элементов четвертой степени M_1^4 множества однородных форм представления знаний M_1 и т. д. Использование комплексных форм представления знаний имеет большое практическое значение. Это особенно важно для повышения эффективности учебного процесса. Используя одновременно несколько однородных форм представления знаний во время чтения лекций, например, можно значительно увеличить степень усвоения учебного материала. Обычно на лекционных занятиях используется речевая форма (s – форма) и текстовая форма в виде записей на доске (t – форма). Если к этим традиционным формам представления знаний добавить видео (v – форма) и графическую форму (g – форма), то это, несомненно, будет способствовать повышению эффективности учебного процесса.

Таким образом, на основании изложенного выше можно предположить с достаточно большой степенью уверенности, что использование комплексных форм представления знаний повышает эффективность учебного процесса.

Различают следующие виды семантической информации (знаний) – первичную и вторичную. Первичной считается семантическая информация, представленная в виде знаков и отображающая сведения о различных сторонах, свойствах и отношениях изучаемого объекта. Это могут быть результаты наблюдения, изучения, исследования изучаемого объекта, а также обобщения, имеющие заверченный характер. При этом форма представления знаний не имеет значения. Адекватной первичной семантической информацией по отношению к объекту познания, считается

информация, которая носит объективный характер и наиболее полно характеризует свойства изучаемого объекта.

Вторичная семантическая информация любой формы представления – это результат аналитико-синтетической и логической обработки первичной семантической информации. Для получения вторичной семантической информации необходим анализ и переосмысление первичной семантической информации, и этот процесс можно рассматривать как продолжение процесса познания объекта, когда знания об объекте становятся более обобщенными и структурированными. Вторичная семантическая информация (вторичные знания об объекте) является моделью первичной семантической информации, которая, в свою очередь, является моделью объекта изучения. Поэтому вторичную семантическую информацию можно рассматривать как модель модели.

Степень толерантности первичных и вторичных знаний об объекте зависит от целей, которые ставятся при обработке первичных знаний. Так, например, степень толерантности реферата на статью выше степени толерантности аннотации, так как реферат еще может сохранять структуру первичного документа, а в аннотации на документ этого сделать невозможно. Носителями первичной семантической информации являются первичные документы. Примерами первичных документов однородной формы представления знаний являются: t – форма (книги, журналы, статьи, диссертации, авторские свидетельства и т. д.), s – форма (аудиозаписи речей, выступлений, музыкальных произведений, стихотворений и т. д.), v – форма (видеозаписи анимации, балета, пантомимы, мимики и жестов), g – форма (картины, рисунки, графики, схемы, карты и т. д.). Вторичными документами однородных форм представления знаний являются: t – форма (рефераты, аннотации, рецензии, карточки библиотечных каталогов и т. д.), s – форма (аудиозаписи фрагментов речей, выступлений, фрагментов музыкальных произведений, и т. д.), v – форма (видеозаписи фрагментов анимации, балета, пантомимы, мимики и жестов), g – форма (обобщенные схемы, контурные карты, приближенные и обобщенные графики, схемы, и т. д.).

Примером первичных знаний комбинированных форм представления может быть кинофильм с субтитрами ($ssvt$ – форма), где s – речь, s – музыка, v – движения актеров, t – текст субтитров. Примерами вторичных знаний комбинированной формы представления информации являются кинореклама, видео клипы, радиореклама, фотомонтаж. В интеллектуальных информационных системах носителями первичных и вторичных знаний всех форм представления являются как человек, так и ЭВМ. Исследования в области интеллектуальных информационных систем позволили сформулировать принцип семантической топологии [3]. Формулируется он так: семантическая информация остается неизменной независимо от формы ее представления. На языке исчисления предикатов этот принцип формулируется следующим образом:

$$\forall f(f \in F) (IS_{f_1} \sim IS_{f_2} \sim \dots \sim IS_{f_i} \sim \dots \sim IS_{f_n}). \quad (7)$$

Запись (7) читается так: для всех форм f представления семантической информации, принадлежащих множеству F форм представления информации, семантическая информация IS_{f_1} формы f_1 эквивалентна семантической информации IS_{f_2} формы f_2 , и т. д. В процессе функционирования интеллектуальных информационных систем этот принцип используется в виде возможности преобразования семантической информации из одной формы представления в другую без изменения ее смысла. Так, например, возможно преобразование аудиальной (s – формы) представления семантической информации в текстовую (t – форму). Уже сейчас существуют так называемые системы речевого ввода, основанные на средствах распознавания речи.

Возможно и обратное преобразование текстовой формы в аудиальную с помощью систем речевого вывода. Аналогичным образом и человек способен осуществлять подобные преобразования при чтении вслух текстовой информации ($t - s$ преобразование) и при записи речевых сообщений ($s - t$ преобразование). Однако в настоящее время не существует единой теории преобразования как первичной, так и вторичной семантической информации. Такие преобразования являются прерогативой человека и осуществляются им интуитивно эвристическими способами. В то же время существует настоятельная необходимость создания искусственных систем преобразования семантической информации из одной формы в другую. Для решения этой очень важной задачи, с нашей точки зрения, необходимо прежде всего разработать математические модели и системы понимания как текстовой (t), так и аудиальной (s) формы представления информации.

Примеры систем $t - s$ и $s - t$ преобразователей приведены на рис. 4 а, б, соответственно, где Y_t – текстовая форма представления семантической информации, а Y_s – аудиальная форма представления семантической информации. Для создания этих систем необходимо разработать математические модели процессов $t - s$ и $s - t$ преобразования на языке алгебры предикатов в виде системы предикатных уравнений и осуществить схемную реализацию полученных моделей. В отличие от существующих систем, которые осуществляют преобразование информации из одной формы в другую программным (алгоритмическим) путем предлагаемые системы основаны на параллельной обработке информации с использованием переключательных схем, построенных на основе разработанных математических моделей.

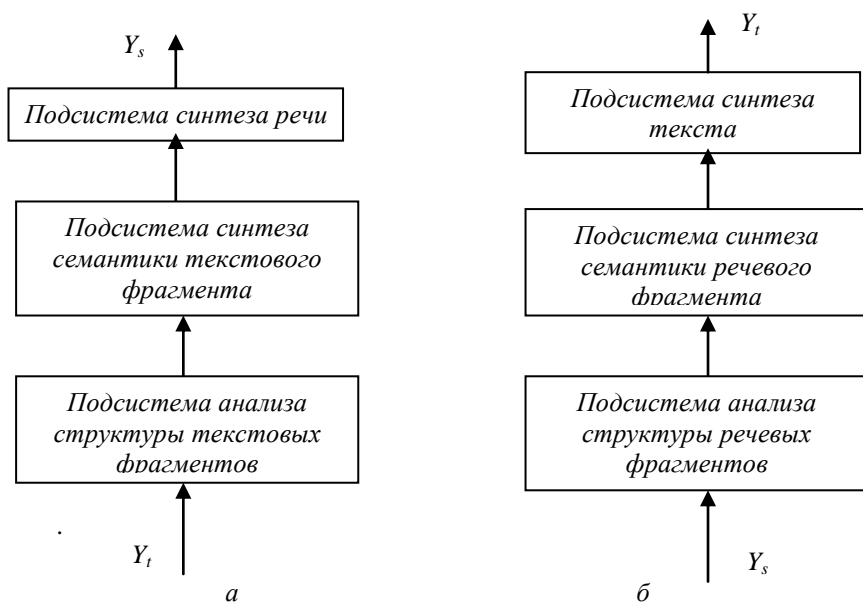


Рис. 4 Текстовая и аудиальная форма представления семантической информации

Значение полученных результатов и выводы. В результате анализа состояния вопросов, связанных с множествами однородных и комплексных форм представления семантической информации, с первичной и вторичной семантической информацией, установлено, что

1) множества однородных и комплексных форм представления информации могут быть заданы в виде уравнений алгебры конечных предикатов;

2) полученные уравнения допускают схемную реализацию в виде переключательных цепей.

Полученные результаты исследований позволили сформулировать принцип семантической топологии; описать возможные методы создания систем преобразования семантической информации из одной формы в другую.

Список литературы: 1. Шабанов –Кушнаренко Ю.П. Теория интеллекта . Математические средства. Х.: ХГУ, 1984. – 144 с. 2. Шабанов –Кушнаренко Ю.П. Теория интеллекта . Технические средства. Х.: ХГУ, 1986. –136 с. 3. Кокорева Л.В., Перевозчикова О.Л. Ющенко Е. Л. Диалоговые системы и представление знаний. Киев: Наукова думка, 1993. – 448 с.

Поступила в редколлегию 00.02.05